Communications for Statistical Applications and Methods
2013, Vol. 20, No. 1, 1–8
DOI: http://dx.doi.org/10.5351/CSAM.2013.20.1.001

Bootstrap aggregated classification for sparse functional data

Hyunsung Kim[a], Yaeji Lim[1,b]

[a]Department of Statistics, Chung-Ang University, Korea
[b]Department of Applied Statistics, Chung-Ang University, Korea

Abstract

Abstract

Keywords: functional principal component analysis, bootstrap aggregating, classification, sparse data

1. Introduction

Prior studies

Description of sections

2. Preliminaries

Functional data analysis (FDA) is a kind of statistics that analyzes the curves or functions rather than single points. In FDA, data (or curve) is defined on the infinite dimension, so dimensionality reduction becomes a key issue. One of the popular dimension reduction method in FDA is functional principal component analysis (FPCA) which finds directions of the variation and exploits by data-driven basis called functional principal component (FPC) scores.

2.1. Functional principal compnent analysis

Let $X(t)$, defined on finite $\mathcal{T}$, is a square integrable random process which means $X(\cdot) \in L_2(\mathcal{T})$ with mean function $\mu(t) = E[X(t)]$ and covariance function $G(s,t) = \mathrm{cov}[X(s), X(t)]$ for $s,t \in \mathcal{T}$. By Mercer's theorem, the covariance function can be represented as $G(s,t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s)\phi_k(t)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ are non-negative eigenvalues satisfying $\sum_{k=1}^{\infty} \lambda_k < \infty$ and ϕ_k's are the corresponding orthonormal eigenfunctions. By the K-truncated Karhunen-Loève expansion, the ith random curve can be represented as

$$X_i(t) = \mu(t) + \sum_{k=1}^{K} \xi_{ik}\phi_k(t), \qquad t \in \mathcal{T}$$

where K is the number of basis functions, and $\xi_{ik} = \int_{\mathcal{T}} (X_i(t) - \mu(t))\phi_k(t)dt$ are uncorrelated variables with mean 0 and variance λ_k.

In the real world, each curve is often observed with measurement errors. Then the ith curve with random noise is denoted as

$$U_i(t) = X_i(t) + \epsilon_i(t), \qquad t \in \mathcal{T}$$

[1] Corresponding author: Professor, Department of Applied Statistics, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Korea. E-mail: yaeji.lim@gmail.com

where $\epsilon_i(t)$ are the uncorrelated measurement errors with mean 0 and variance σ^2.

The number of basis functions, K, is often selected by proportion of variance explained (PVE). Yao *et al.* (2005) presented AIC that is more efficient in computation than cross-validation. Li *et al.* (2013) proposed BIC and proved its consistency on FPCA.

2.2. Functional principal component analysis for sparse functional data

When each curve is observed at sparse or irregular time points, we can't apply above method directly. The covariance function can't be computed easily and also estimated FPC scores by numerical integration be biased. So, James *et al.* (2001) proposed the reduced rank model based on the mixed effects model and estimate FPC function and scores by EM algorithm. Yao *et al.* (2005) proposed principal component analysis through conditional expectation (PACE) method to obtain the unbiased FPC scores.

Let $t_{ij} \in \mathcal{T}$ is the jth time point observed in the ith curve $X_i(\cdot)$ where $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n_i$. Then, the ith curve for sparse functional data using from K-truncated FPCA can be expressed as

$$U_i(t_{ij}) = \mu(t_{ij}) + \sum_{k=1}^{K} \xi_{ik}\phi_k(t_{ij}) + \epsilon_i(t_{ij}), \qquad t_{ij} \in \mathcal{T}$$

where K is the number of basis functions, and $\epsilon_i(t_{ij})$ are the random noises.

Let ith curve $\mathbf{X}_i = (X_i(t_{i1}), \ldots, X_i(t_{in_i}))^T$, ith curve with measurment error $\mathbf{U}_i = (U_i(t_{in_1}), \ldots, U_i(t_{in_i}))^T$, the mean function $\boldsymbol{\mu}_i = (\mu(t_{i1}), \ldots, \mu(t_{in_i}))^T$, the FPC functions $\boldsymbol{\phi}_{ik} = (\phi_k(t_{i1}), \ldots, \phi_k(t_{in_i}))^T$, and the measurment errors $\boldsymbol{\epsilon}_i = (\epsilon_i(t_{i1}), \ldots, \epsilon_i(t_{in_i}))^T$. The best linear unbiased prediction (BLUP) of ξ_{ik}, kth FPC scores of ith subject, by PACE is

$$\tilde{\xi}_{ik} = E[\xi_{ik}|\mathbf{U}_i] = \lambda_k \boldsymbol{\phi}_{ik}^T \boldsymbol{\Sigma}_{\mathbf{U}_i}^{-1}(\mathbf{U}_i - \boldsymbol{\mu}_i),$$

where $\boldsymbol{\Sigma}_{\mathbf{U}_i} = \text{cov}(\mathbf{U}_i, \mathbf{U}_i) = \text{cov}(\mathbf{X}_i, \mathbf{X}_i) + \sigma^2 \mathbf{I}_{n_i}$,

From the above, the PACE estimate of ξ_{ik} is obtained as follow

$$\hat{\xi}_{ik} = \widehat{E}[\xi_{ik}|\mathbf{U}_i] = \hat{\lambda}_k \hat{\boldsymbol{\phi}}_{ik}^T \widehat{\boldsymbol{\Sigma}}_{\mathbf{U}_i}^{-1}(\mathbf{U}_i - \hat{\boldsymbol{\mu}}_i).$$

where $\widehat{\boldsymbol{\Sigma}}_{\mathbf{U}_i} = \widehat{\text{cov}}(\mathbf{U}_i, \mathbf{U}_i) = \widehat{\text{cov}}(\mathbf{X}_i, \mathbf{X}_i) + \hat{\sigma}^2 \mathbf{I}_{n_i}$.

How to apply FPCA for sparse data?

3. Method (Functional classifier ensemble)

3.1. Functional classification(Classification based on functional principal component scores)

For classification, functional generalized linear model (FGLM) was proposed by James (2002) and Müller & Stadtmüller (2005).

Let $\mathbf{X}_i = X_i(\cdot)$ is the ith curve, Y_i is its response, $\beta(\cdot)$ is the coefficient function, and $g(\cdot)$ is a link function. Denote $\mu = E(Y_i|\mathbf{X}_i)$, then, the functional generalized linear model (FGLM) is

$$\begin{aligned} g(\mu) &= \alpha + \int_{\mathcal{T}} \beta(t) X_i(t) dt \\ &= \alpha + \sum_{t \in \mathcal{T}} \beta(t) X_i(t). \end{aligned}$$

Since we can only observe $X_i(t)$ at a finite time points, the integral can be substituted with a summation. But it has a problem that the estimate of $\beta(\cdot)$ may be unstable since $\beta(\cdot)$ be an extremely high dimensional vector.

We can solve this problem by expanding $\beta(\cdot)$ in terms of a set of basis functions. If FPCA is performed, we can also use data-driven orthonormal basis functions to represent $\beta(\cdot)$. In addition, it is stable because most of the variation can be expressed in a small number, K.

Denote $X_i(t) = \sum_{k=1}^{K} \xi_{ik}\phi_k(t)$ and $\beta(t) = \sum_{k=1}^{K} \beta_k\phi_k(t)$ from K-truncated FPCA model, then the FGLM is represented as

$$g(\mu) = \alpha + \sum_{k=1}^{K} \beta_k \xi_{ik}.$$

If we set $g(\cdot)$ as logit link, the FGLM above becomes functional logistic regression for binary classification problems. Therefore, we can construct the functional classifier by modeling a conventional classifier on FPC scores.

3.2. Bootstrap aggregating

Bootstrap aggregating (Bagging) is the ensemble method using bootstrap ideas proposed by Breiman (1996). It is a method of extracting samples several times with replacement and learning each model to aggregate results. It can avoid overfitting problems and also reduce its variance. Generally, it is aggregated into an average in regression problems, and a majority vote in classification problems.

3.2.1. Majority vote

Suppose there are g response classes with $C_1, \dots, C_g$, and a classifier, $\hat{f}^{(b)}(x)$ for $b = 1, \dots, B$, obtained from bth bootstrap resample. Then, the estimate of bagged classifier by the majority vote is

$$\hat{y}_{\text{bag}} = \arg\max_j \#\{b \mid \hat{f}^{(b)}(x) = C_j\}.$$

3.2.2. Out-of-bag accuracy weighted vote

3.3. Bootstrap aggregated functional classifier with sparse FPCA

We propose the method that aggregate independent functional classifiers from bootstrap resamples. For dimension reduction, FPCA with PACE is performed for each functional classifier.

Suppose we have n curves, $\mathbf{U}_1, \dots, \mathbf{U}_n$ observed at sparse and irregular time points. Also, we have the response variable $y_1, \dots, y_n$ with g different class labels for each curve. Denote the set of sparse n curves as $\mathcal{D} = \{(\mathbf{U}_i, y_i) \mid i = 1, \dots, n\}$ and $\mathcal{D}^{(b)} = \left\{\left(\mathbf{U}_i^{(b)}, y_i^{(b)}\right) \mid i = 1, \dots, n\right\}$ for $b = 1, \dots, B$ is a bootstrap resample from $\mathcal{D}$. For each bootstrap functional data, we perform FPCA with PACE approach. First K FPC scores are chosen by the appropriate criterian, PVE, BIC, cross-validation, etc. In this paper, we select K satisfiying PVE is greater than 0.99. Using these K FPC scores, we construct a classifier, like support vector machine (SVM), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), etc. Then, we can get B classifiers from each bootstrap sample. When we observe a new curve, we can obtain B predictions from B classifiers and aggregate it by majority vote. Then, we can get the prediction from bagged classifier with sparse FPCA. The summary of procedure is shown in the below algorithm.

Algorithm 1: Bagged functional classifier with sparse FPCA

1. For $b = 1, \ldots, B$, repeat
   (a) Generate a bootstrap resample $\mathcal{D}^{(b)}$ from the data $\mathcal{D}$.
   (b) Perform functional principal component analysis for $\mathcal{D}^{(b)}$.
   (c) Estimate the FPC scores by PACE.
   (d) Select K, the number of FPCs, such that PVE ≥ 0.99.
   (e) Construct a classifier on K FPC scores.
2. Given a new curve $\mathbf{U}^*$, for $b = 1, \ldots, B$, repeat
   (a) Estimate the FPC scores by PACE using FPC function from $\mathcal{D}_b$.
   (b) Obtain the prediction $\hat{y}^{(b)}$ from above bth classifiers.
3. From $\hat{y}^{(1)}, \ldots, \hat{y}^{(B)}$, obtain the final prediction $\hat{y}_{bag}$ aggreated by majority vote.

algorithm flow image??

4. Simulation studies

4.1. Simulation 1

In this simulation, we consider 3 situations, (A) different mean and variance, (B) different mean, and (C) different variance.

We generate a total of $n = 200$ curves with 2 classes ($n/2$ for each class) from $U_{gi}(t) = \mu_g(t) + \sum_{k=1}^{3} \xi_{gk}\phi_k(t) + \epsilon_i(t)$ for $g = 0, 1$ and $t \in [0, 10]$. The FPC functions are $\phi_k(t) = \cos(\pi kt/5)/\sqrt{5}$ for k is odds and $\phi_k(t) = \sin(\pi kt/5)/\sqrt{5}$ for k is even. In Table 1, the parameters of mean and variance are decided for each models. The mean functions $\mu_g(t)$ are generated for each class and the FPC scores ξ_{gk} are sampled from $i.i.d.\ N(0, \lambda_{gk})$ for $k = 1, 2, 3$. The measurement error $\epsilon_i(t)$ is sampled from $i.i.d.\ N(0, 0.5^2)$ To make each curve sparse, the number of grids for ith curve n_i is randomly selected from 5 to 10, and the observed time points t_{ij}, $j = 1, \ldots, n_i$, are selected from $i.i.d.\ Uniform(0, 10)$.

To compare the proposed method with single functional classifier, we randomly split the generated data to training and test set with 100 each. We consider 7 classification models (logistic regression, SVM with linear kernel, SVM with gaussian kernel, LDA, QDA and naive bayes), and the truncation number K is selected to satisfy PVE ≥ 0.99 in FPCA. We compute percentage of classification error and standard error for 100 Monte Carlo repetitions for each classifier. 앞에서 SVM, LDA, QDA 언급 필요

Table 1: The parameters of 3 different models from simulation 1

Model	g	$\mu_g(t)$	λ_g
A	0	$t + \sin(t)$	$(4, 2, 1)$
	1	$t + \cos(t)$	$(16, 8, 4)$
B	0	$t + \sin(t)$	$(4, 2, 1)$
	1	$t + \cos(t)$	$(4, 2, 1)$
C	0	$t + \sin(t)$	$(4, 2, 1)$
	1	$t + \sin(t)$	$(16, 8, 4)$

Shown in Table 2, 결과 언급

Table 2: The average classification error with standard error in percentage from 100 Monte Carlo repetitions for 3 different simulation designs with different class proportion

Model	$P(C_1)$	Method	Logistic Regression	SVM (Linear)	SVM (Gaussian)	LDA	QDA	Naive Bayes
A	0.5	Single	17.60 (4.84)	17.48 (5.12)	15.28 (5.30)	17.07 (4.74)	15.07 (4.82)	16.48 (4.55)
		Majority vote	15.47 (4.23)	15.66 (4.41)	13.16 (4.35)	15.53 (4.08)	13.63 (4.17)	15.23 (4.09)
		OOB weight	16.14 (4.25)	16.32 (4.49)	13.85 (4.34)	16.30 (4.17)	14.20 (4.12)	15.71 (3.97)
	0.4	Single	17.65 (4.91)	17.71 (5.27)	15.75 (5.84)	17.72 (4.98)	15.19 (4.82)	16.31 (4.78)
		Majority vote	15.68 (4.02)	15.56 (4.00)	13.74 (4.58)	15.39 (3.97)	14.50 (4.68)	15.51 (4.44)
		OOB weight	16.24 (3.77)	16.33 (3.90)	14.49 (4.22)	16.13 (3.85)	15.07 (4.57)	15.95 (4.41)
	0.3	Single	17.26 (5.50)	17.87 (6.12)	15.21 (5.57)	17.17 (5.32)	14.89 (5.24)	15.23 (5.04)
		Majority vote	15.90 (4.35)	16.40 (5.31)	13.25 (4.67)	15.97 (4.17)	14.64 (4.24)	14.98 (4.14)
		OOB weight	16.71 (4.19)	17.34 (5.14)	14.15 (4.68)	16.87 (3.95)	15.37 (4.23)	15.72 (4.38)
	0.2	Single	15.60 (4.31)	16.31 (4.31)	14.32 (4.92)	15.45 (4.31)	13.55 (4.51)	13.91 (4.12)
		Majority vote	14.19 (4.14)	15.12 (4.70)	12.80 (4.61)	14.04 (3.91)	13.92 (3.88)	13.70 (3.99)
		OOB weight	15.26 (4.20)	16.31 (5.00)	14.01 (4.67)	15.08 (3.96)	15.02 (4.03)	14.81 (4.03)
B	0.5	Single	11.88 (3.43)	11.45 (3.49)	12.00 (4.25)	11.25 (3.55)	12.85 (3.62)	13.97 (4.39)
		Majority vote	10.70 (3.29)	10.44 (3.13)	10.96 (3.70)	10.44 (3.24)	11.59 (3.36)	12.41 (3.38)
		OOB weight	11.39 (3.27)	11.21 (3.00)	11.75 (3.60)	11.05 (3.30)	12.32 (3.36)	13.07 (3.59)
	0.4	Single	11.13 (3.49)	10.99 (3.38)	11.93 (3.78)	10.84 (3.4)	12.11 (3.62)	13.72 (4.21)
		Majority vote	9.96 (2.88)	9.81 (2.58)	10.73 (3.25)	9.74 (2.79)	10.82 (3.15)	11.95 (3.33)
		OOB weight	10.85 (2.92)	10.64 (2.79)	11.53 (3.27)	10.63 (2.92)	11.74 (3.26)	12.73 (3.56)
	0.3	Single	10.23 (3.55)	10.28 (3.51)	11.10 (3.77)	9.62 (3.22)	11.33 (3.44)	12.33 (3.43)
		Majority vote	9.35 (3.29)	9.39 (3.16)	9.62 (2.95)	9.07 (2.82)	10.31 (3.01)	10.93 (3.01)
		OOB weight	10.36 (3.46)	10.45 (3.33)	10.69 (3.26)	10.17 (2.99)	11.29 (3.27)	12.00 (3.21)
	0.2	Single	9.61 (2.98)	9.14 (2.79)	9.54 (3.65)	8.82 (2.89)	10.63 (3.86)	10.65 (3.43)
		Majority vote	8.41 (2.68)	8.21 (2.46)	8.30 (2.66)	7.81 (2.55)	9.37 (3.07)	9.30 (2.85)
		OOB weight	9.67 (2.93)	9.57 (2.74)	9.61 (2.87)	9.15 (2.86)	10.60 (3.31)	10.64 (3.15)
C	0.5	Single	50.49 (5.65)	49.53 (5.47)	32.79 (5.03)	50.59 (5.63)	31.20 (4.51)	30.50 (4.71)
		Majority vote	49.45 (5.58)	48.25 (6.15)	31.27 (5.09)	49.50 (5.68)	30.76 (4.06)	29.80 (4.29)
		OOB weight	48.85 (5.53)	47.83 (6.03)	31.51 (5.13)	48.81 (5.47)	31.01 (4.03)	30.11 (4.25)
	0.4	Single	47.78 (4.25)	41.04 (4.32)	33.24 (5.06)	47.55 (4.17)	31.12 (4.61)	30.04 (4.47)
		Majority vote	47.23 (4.39)	40.78 (3.51)	32.71 (4.90)	47.14 (4.54)	30.71 (4.25)	29.42 (4.04)
		OOB weight	47.56 (4.65)	41.12 (3.63)	33.22 (4.76)	47.43 (4.59)	31.06 (4.08)	29.78 (4.01)
	0.3	Single	33.58 (3.97)	30.09 (3.36)	29.61 (4.00)	33.26 (3.77)	27.49 (4.33)	26.85 (4.01)
		Majority vote	33.19 (4.02)	30.09 (3.36)	28.96 (3.91)	32.91 (3.78)	26.95 (3.87)	26.44 (3.82)
		OOB weight	33.93 (4.20)	30.79 (3.51)	29.78 (4.00)	33.62 (3.94)	27.91 (3.85)	27.19 (3.61)
	0.2	Single	21.06 (2.91)	19.65 (2.67)	20.26 (3.19)	20.86 (2.75)	21.31 (3.45)	21.01 (3.2)
		Majority vote	20.64 (2.78)	19.65 (2.67)	19.93 (2.89)	20.54 (2.85)	20.09 (2.85)	20.52 (2.65)
		OOB weight	21.79 (3.12)	20.77 (3.00)	21.10 (3.23)	21.69 (3.19)	21.42 (3.16)	21.77 (3.01)

OOB weight = Out-of-bag weighted vote

4.2. Simulation 2

In second simulation, we refer the procedure of generating data from Yao *et al.* (2016).

$$\text{Model A: } f(x) = \exp(\langle \beta_1, X \rangle / 2) - 1,$$
$$\text{Model B: } f(x) = \arctan(\pi \langle \beta_1, X \rangle) + \exp(\langle \beta_2, X \rangle / 3) - 1,$$
$$\text{Model C: } f(x) = \arctan(\pi \langle \beta_1, X \rangle / 4).$$

Shown in Table 3, 결과 언급

Data generating

100 Monte Carlo simulation(generate + split)
Result

Table 3: The average classification error with standard error in percentage from 100 Monte Carlo repetitions for 3 different simulation designs

Model	Method	Logistic Regression	SVM (Linear)	SVM (Gaussian)	LDA	QDA	Naive Bayes
A	Single	16.71 (2.33)	16.82 (2.20)	17.50 (2.76)	16.62 (2.30)	17.77 (2.56)	18.41 (2.66)
	Majority vote	15.62 (1.95)	15.86 (1.87)	16.19 (2.28)	15.79 (1.96)	16.51 (2.14)	17.32 (2.42)
	OOB weight	16.00 (2.02)	16.21 (1.94)	16.56 (2.28)	16.14 (1.98)	16.86 (2.09)	17.71 (2.43)
B	Single	12.85 (2.41)	12.79 (2.40)	13.27 (2.65)	12.77 (2.40)	13.83 (2.56)	14.77 (2.74)
	Majority vote	11.20 (1.84)	11.14 (1.89)	11.54 (1.98)	11.19 (1.85)	11.93 (2.03)	13.29 (2.36)
	OOB weight	11.62 (1.86)	11.54 (1.90)	11.96 (1.96)	11.59 (1.86)	12.33 (2.06)	13.63 (2.35)
C	Single	14.46 (2.17)	14.34 (2.18)	15.27 (2.69)	14.29 (2.17)	15.32 (2.36)	16.05 (2.22)
	Majority vote	13.15 (1.73)	13.14 (1.78)	13.62 (2.08)	13.14 (1.82)	13.78 (1.90)	14.88 (2.09)
	OOB weight	13.53 (1.81)	13.50 (1.78)	13.99 (2.03)	13.52 (1.84)	14.18 (1.92)	15.22 (2.12)

OOB weight = Out-of-bag weighted vote

5. Real data analysis

5.1. Berkely growth data

To compare the proposed method with single classifier, we consider the berkely growth data presented in Tuddenham & Snyder (1954). The dataset was measured height for 93 individuals, 54 girls and 39 boys. There are 31 observations from ages 1 to 18 for each curve. These dense curves with 54 girls and 39 boys are shown in Figure 1.

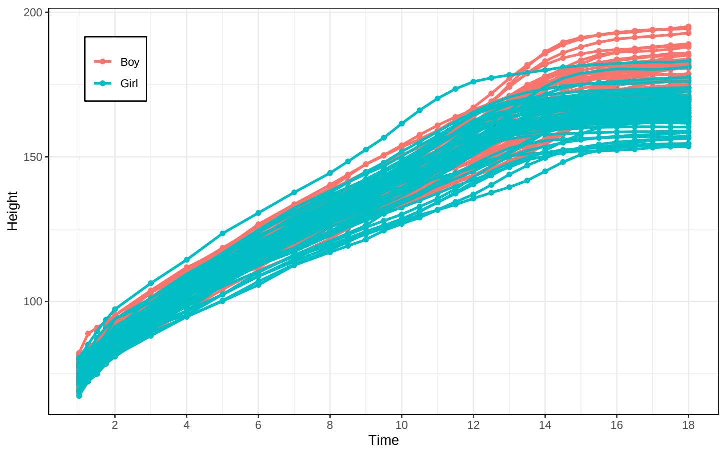

Figure 1: *The berkely growth data of 93 individuals.*

Our method is to apply to sparse data, so we have transformed data into arbitrarily sparse. The number of observation n_i was randomly selected from 12 to 15 and its time points also randomly selected from original time points. We randomly selected 62 curves as training set and the remaining 31 curves as test set. For these sparse data, we performed the proposed bagged functional classifiers with sparse FPCA and compared to the single classifiers. The total of 6 models, logistic regression, SVM with linear kernel, SVM with gaussian kernel, LDA, QDA and naive bayes were performed. We repeated this process 100 times for different split. There are the average classification error and its standard error in Table 4. All bagged classifiers showed better performances than single models, and bagged QDA showed the best.

Table 4: The average classification error with standard error in percentage from 100 random split for berkerly growth data

Method	Logistic Regression	SVM (Linear)	SVM (Gaussian)	LDA	QDA	Naive Bayes
Single	7.3 (4.80)	5.3 (3.20)	5.7 (4.03)	5.8 (3.34)	5.6 (3.35)	5.6 (3.90)
Majority vote	5.9 (4.12)	4.9 (3.19)	5.3 (3.51)	5.4 (3.24)	4.9 (3.57)	5.5 (3.96)
OOB weight	5.9 (4.12)	5.0 (3.22)	5.4 (3.62)	5.4 (3.27)	4.9 (3.54)	5.5 (3.96)

Data description
Result

5.2. Spinal bone mineral density data

Second, we consider the sparse functional data, the spinal bone mineral density data presented in Bachrach *et al.* (1999). The dataset has the spinal bone mineral density for 280 individuals, 153 females and 127 males measured at sparse and irregualar time points. There are $2 \sim 4$ observations for each curve. The dataset has also ethnicity information such that Asian, Black, Hispanic and White. For this data, we consider gender classification. These sparse curves with 153 females and 127 males are shown in Figure 2.

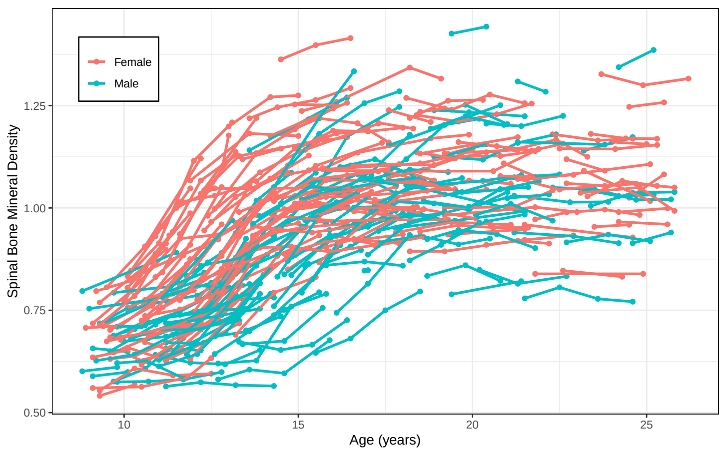

Figure 2: *The spinal bone mineral density of 280 individuals.*

We performed the same procedure except to make the data sparse. The data was randomly divided into 187 training and 93 test sets. The same 6 classifiers were considered on the above and both single and bagged classifiers were performed for each classifier. This procedures were applied to 100 different splits. The results are shown in Table 5. The proposed methods showed better performance than single classifiers. The bagged logistic regression with majority vote showed the best performance in this data.

Table 5: The average classification error with standard error in percentage from 100 random split for spinal bone mineral density data

Method	Logistic Regression	SVM (Linear)	SVM (Gaussian)	LDA	QDA	Naive Bayes
Single	31.3 (4.30)	32.0 (4.27)	33.2 (4.71)	31.4 (4.44)	33.3 (4.10)	32.3 (4.33)
Majority vote	30.2 (3.72)	30.8 (4.18)	31.2 (3.88)	30.4 (3.77)	31.6 (3.78)	30.9 (3.83)
OOB weight	30.3 (3.71)	30.8 (4.07)	31.4 (3.81)	30.5 (3.82)	31.8 (3.71)	30.9 (3.86)

6. Conclusion and discussion

conclusion and result

Acknowledgement

This research was supported by the Chung-Ang University Graduate Research Scholarship in 2019.

References

Bachrach, L. K., Hastie, T., Wang, M. C., Narasimhan, B., & Marcus, R. (1999). Bone mineral acquisition in healthy Asian, Hispanic, black, and Caucasian youth: a longitudinal study. *The journal of clinical endocrinology & metabolism*, **84**, 4702-4712.

Breiman, L. (1996). Bagging predictors. *Machine learning*, **24**, 123-140.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning.*, Springer series in statistics, New York.

James, G. M., Hastie, T. J., & Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika*, **87**, 587-602.

James, G. M., & Hastie, T. J. (2001). Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **63**, 533-550.

James, G. M. (2002). Generalized linear models with functional predictors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**, 411-432.

Li, Y., Wang, N., & Carroll, R. J. (2013). Selecting the number of principal components in functional data. *Journal of the American Statistical Association*, **108**, 1284-1294.

Müller, H. G., & Stadtmüller, U. (2005). Generalized functional linear models. *the Annals of Statistics*, **33**, 774-805.

Ramsay, J. O., & Silverman, B. W. *Functional Data Analysis*, 2nd ed., Springer Series in Statistics, New York.

Tuddenham, R. D., & Snyder, M. M. (1954). Physical growth of California boys and girls from birth to eighteen years. *University of California publications in child development*, **1**, 183-364.

Yao, F., Müller, H. G., & Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, **100**, 577-590.

Yao, F., Wu, Y., & Zou, J. (2016). Probability-enhanced effective dimension reduction for classifying sparse functional data. *Test*, **25**, 1-22.

Received 00, 0000; Revised 00, 0000; Accepted 00, 0000